Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment

Iordanou, Kalypso and Constantinou, Costas

Iordanou, Kalypso ORCID: 0000-0001-5930-9393 and Constantinou, Costas (2015) Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. Science Education, 99 (2). pp. 282-311. ISSN 0036-8326

Supporting Use of Evidence in Argumentation through Practice in Argumentation and Reflection in the context of SOCRATES Learning Environment

[a]Kalypso Iordanou and [b]Costas P. Constantinou

[a]University of Central Lancashire, Cyprus

[b]University of Cyprus, Cyprus

Corresponding Author: Kalypso Iordanou,

Email: KIordanou@uclan.ac.uk

University of Central Lancashire, Cyprus

12-14 University Avenue, Pyla, 7080 Larnaka, Cyprus

Abstract

The aim of this study was to examine how students used evidence in argumentation while they engaged in argumentive and reflective activities in the context of a designed learning environment. A web-based learning environment, SOCRATES, was developed, which included a rich data base on the topic of Climate Change. Sixteen 11th graders, working with a partner, engaged in electronic argumentive dialogs with classmates who held an opposing view on the topic and in some evidence-focused reflective activities, based on transcriptions of their dialogs. Another sixteen 11th graders, who studied the data base in the learning environment for the same amount of time as experimental-condition students but did not engage in an argumentive discourse activity, served as a comparison condition. Students who engaged in an evidence-focused dialogic intervention increased the use of evidence in their dialogs, used more evidence that functioned to weaken opponents' claims and used more accurate evidence. Significant gains in evidence use and in meta-level communication about evidence were observed after students engaged in reflective activities. We frame our discussion of these findings in terms of their implications for promoting use of evidence in argumentation, and in relation to the development of epistemological understanding in science.

Engagement in argument from evidence is one of the fundamental objectives of science education from kindergarten through grade 12 (NGSS Lead States, 2013). There is broad consensus on the need to move away from practices that support the mere transmission of facts for assimilation by students, to teaching and learning practices that promote scientific thinking. Argumentation lies at the heart of scientific thinking (Kuhn, 2010; Kuhn, Iordanou, Pease, & Wirkala, 2008). Judging scientific theories and offering alternative views for interpreting data are fundamental skills for scientific thinking. The importance of developing argumentation has been identified both by science education researchers (Duschl, 2008; Erduran & Jimenéz-Aleixandre, 2008) and policy makers (NGSS Lead States, 2013).

A fundamental element of skilled scientific argumentation is evidence. Evidence is considered an essential component of both strong individual argument (Toulmin, 1958) and skilled dialogic argument, namely argumentation.

> In dialogic argument at a minimum one must recognize an opposition between two assertions that, on surface appearance at least, both are not correct. Evidence must then be related to each of the assertions, and, ideally, if the argument is to move toward resolution, this evidence needs to be weighed in an integrative evaluation of the relative merits of the opposing assertions. (Kuhn, 1991, p. 12)

Despite its importance, research has shown that students struggle with scientific argumentation (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kelly & Takao, 2002; Kolstø et al., 2006). Little is known about how we can support students in developing the ability to engage in scientific argumentation, particularly their ability to employ evidence to support their claims and critiques they might offer to an opponent’s claims. The present study sought to examine how students used evidence in argumentation while they were engaging in argumentive and reflective activities in the context of a web-based learning environment. Another group of high school students who studied the same information as experimental-condition students but did not engage in argumentive and reflective activities served as a comparison condition.

Background

Research on argumentation has been flourishing lately both inside and outside the science education domain, with researchers using quite different perspectives to approach argumentation (Erduran & Jiménez-Aleixandre, 2008). In the work presented here, we share the view that the development of scientific argumentation is multi-faceted (Duschl, 2008; Kuhn, 2010). In particular, we adopt Kuhn's (2010) model of argumentation, according to which argumentation has both procedural and meta-level components that regulate its use. The procedural components involve the cognitive skills that support the execution of argumentation, while the meta-level components involve both meta-strategic understanding of the goals of argumentation and more general epistemological understanding, that is, understanding of what is scientific knowledge and how one knows.

Research offers cumulative evidence showing that students struggle with scientific argumentation (Jimenez-Aleixandre et al., 2000; Kelly & Takao, 2002; Kolstø et al., 2006). We believe that underlying these struggles are insufficient developments at the meta-level understanding, meta-strategic or epistemological, that support argumentation. According to Walton (1989) argumentation has two goals. The first is to secure commitments from the opponent that can be used to support one's own argument. The second is to undermine the opponent's position by identifying and challenging weaknesses in the opponent's argument. Without approaching argumentation through Walton's lenses, evaluating and critiquing others' ideas would appear meaningless. Berland and Reiser (2011) showed that students who were asked to argue with a persuasive goal engaged in evaluation and critique of others' ideas more frequently than students who were asked to argue with a goal of sense-making, providing evidence for a relationship between students' goals in argumentation and the employment of specific argumentative moves. According to Berland and Reiser,

> unless students genuinely took on the goal of trying to persuade others, who did not already know what they knew—why their claim is supported by evidence—there is little motivation for them to go beyond presenting the story they thought correct. (2009, p. 48)

Furthermore, previous research has shown that students who haven't developed a constructive epistemological understanding do not engage in skilled argumentation (Mason & Scirica, 2006; Nussbaum, Sinatra, & Poliquin, 2008). For example, the work of Nussbaum, Sinatra, and Poliquin (2008) showed that students who haven't developed a constructive epistemological understanding interacted less critically when engaged in argumentation than students who had done so.

Various approaches have been developed to help students learn how to participate in scientific argumentation with mixed results **(**Osborne, Erduran & Simon, 2004; Zohar & Nemet, 2002)**.** Some efforts to support argumentation focused on scaffolding students' understanding of the structure of a "good" argument based on Toulmin's argumentation model, with the objective to make explicit the importance of making claims that can be justified with scientific evidence (Erduran, Simon, & Osborne, 2004; Krajcik, & Marx, 2006; McNeill, Lizotte, Sampson & Clark, 2009; Sandoval & Reiser, 2004; Zohar & Nemet, 2002). These efforts have shown that scaffolds can help students construct written scientific explanations (McNeil et al., 2006; Sampson & Clark, 2009), produce individual written arguments (Zohar & Nemet, 2002) and develop conceptual understanding on a particular topic (Bell & Linn, 2000; McNeil, Pimentel., & Strauss, 2013; Zohar & Nemet, 2002). Yet, the modest gains of argumentation-supported instruction (Osborne, Erduran & Simon, 2004; Zohar & Nemet, 2002) point to the challenge of supporting the development of scientific argumentation and to the need for further research in order to gain a better understanding of how to support students in their development of the ability to engage in skilled argumentation.

Other efforts to support argumentation have focused on offering professional development to teachers (Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013) and studying teachers' practices in order to gain an understanding of what kind of practices support scientific argumentation (McNeill, Pimentela, & Strauss, 2013; Ryu & Sandoval, 2012). For example, McNeill, Pimentela, and Strauss (2013) observed teachers while using an ecology curriculum and found that teachers who spent a larger percentage of time on group work and offered students the opportunity to engage in argumentation instead of attending lectures had greater success in supporting the development of students' understanding of specific science concepts and their ability to apply these concepts when constructing written arguments. Similarly, Ryu and Sandoval (2012), in observing the classroom of an "exceptional" teacher over an academic year, found that engagement in group work along with scaffolding offered by the teacher, in the form of reflective questions, promoted students' ability to use evidence and offer justification in written arguments. Yet, the work of Osborne et al. (2013) showed that relying solely on teachers, without any particular curriculum, is not always a successful means to promote students' argumentation. In their study, Osborne and his colleagues offered a 5-day professional training focused on encouraging teachers to engage students in argumentation. Then the trained teachers supported their colleagues in using argumentative activities in their teaching practice. Pre- and post-intervention evaluations showed that students whose teachers participated in the intervention did not exhibit any considerable improvements in their argumentation skills. Furthermore, the work of McNeil (2009) showed that even when teachers received a specific curriculum to support students in writing scientific arguments, this was not a sufficient condition for supporting students' scientific argumentation. McNeil (2009) showed that there was great variability in the way that teachers used a particular curriculum, which resulted in variability in the subsequent student gains in terms of their ability to write scientific arguments to explain phenomena. The findings of research examining teachers as facilitators for supporting students' scientific argumentation show that offering teachers pedagogical guidelines that encourage them to use argumentation or even providing teachers with a specific curriculum is not always a sufficient condition to promote students' scientific argumentation (McNeil et al., 2013; Osborne et al., 2013). Teachers' beliefs and views about argumentation play a determining role in whether and how they support students in their development scientific argumentation. Besides supporting teacher development of an appreciation of the value of having students engaged in argumentation, the development of a technology-enhanced curriculum that will be more student-centered and relies less on teachers for providing scaffolding might be a promising way for supporting students in developing scientific argumentation.

In the present study, we examine whether a student-centered curriculum involving engagement in a series of dialogs with peers holding an opposing position on a socio-scientific topic along with reflective activities on the dialogs produced, in the context of a web-based learning environment, can support scientific argumentation. We extend previous work that has focused on scaffolding students' understanding of the structure of argument as a product by focusing on the process of argument construction, that is argumentation, to promote skilled argumentation. Our method is based on the view that the two forms of argument—individual and dialogic—are intrinsically connected (Billig, 1987). Individual argument, like dialogic argument, involves the presentation of one's own position along with evidence in support of own position, but also presentation of an opposing position, along with evidence supporting the opposing position, and an implicit weighing process to establish the superiority of one's own position. Hence, individual argument encompasses an implicit dialogic argument. We share the view that dialogic argumentation is a promising pathway for the development of scientific argumentation, because it provides the "missing interlocutor" (Graff, 2003) that is lacking in the individual argument and makes the goal of persuasion meaningful. The presence of real audience that needs to be convinced is a facilitating factor for students to develop an understanding of the need to provide justification and evidence in their arguments, instead of reporting their self-evident, "right" claims (Berland & Reiser, 2009). Another benefit of dialogic argumentation is that it makes thinking visible and this visibility provides a "powerful mediation or formative assessment opportunity" (Duschl, 2008). Our method is also based on the view that skilled argumentation is developed through practice in argumentation (Berland & Reiser, 2009, 2011; Iordanou, 2010; Ryu & Sandoval, 2012; Sandoval, 2005). Our approach of practice is based on the claim underlying micro-genetic research (Kuhn, 1995) that dense exercise of existing strategies over a period of time is a promising condition for change.

Previous work has shown that engagement and practice in dialogic argumentation is a promising method for supporting the development of scientific argumentation (Iordanou, 2010; McNeill, 2009). The work of Iordanou (2010) showed that sixth graders who engaged in dialogic argumentation with classmates who shared an opposing position on the topic of dinosaurs' extinction as well as in some reflective activities based on transcriptions of their dialogs, shifted their attention from presenting one's own position to critiquing opposing position, through counterarguments, in dialogic argumentation. Besides developments at the procedural level, developments at the meta-level have been reported. In particular, the work of Kuhn, Goh, Iordanou and Shaenfield (2008) showed that participants who engaged in dialogic argumentation over an extended period of time showed significant gains in meta-level communications about the discourse, reflecting at least implicit understanding of its goals as well as the strategic moves that constituted the discourse. In addition, gains in students' epistemological understanding have also been reported as a result of participating in argumentation (Iordanou, 2010; Ryu & Sandoval, 2012). In particular, Ryu and Sandoval (2012) found that students who participated in sustained argumentation offered explicit justifications regarding the fit between evidence and claims when asked to evaluate arguments.

The present work examines how students use evidence over time when they engage in an evidence-focused intervention based on engagement and practice in argumentive dialogic activities and in some reflective activities on a socio-scientific topic. Meta-level awareness was facilitated by conducting the dialogs via instant messaging software, which made available a transcript of the dialog subsequently used in additional reflective activities. Arguing on the computer has the benefit of providing an immediately available, permanent record of the discourse for participants to reflect on, in contrast to the conditions of real-time verbal discourse, where the contents of each contribution to the dialog immediately disappear as soon as they are spoken. The intervention method used in the work presented here is modeled on the method of engagement and practice in argumentation using instant-messaging (IM) computer software as the medium of discourse, following the successful use of this method in supporting the development of students' skill in producing counterarguments and rebuttals when arguing on a science topic (Iordanou, 2010). The gains in argumentation skills developed in the context of arguing on the computer were not only evident when students argued electronically but successfully transferred when students argued face-to-face (Iordanou, 2013).

Several studies suggest that the computer is a fruitful medium for scaffolding argumentation in the science domain (see Clark, Stegmann, Weinberger, Menekse, & Erkens, 2008, for a review of studies using technology-enhanced environments to support argumentation). For example, Bell and Linn (2000) developed the KIE learning environment for scaffolding students in using evidence in their arguments. KIE supported students' use of evidence from the web to develop a written argument by providing hints and models of arguments from experts. The present study shares some features that Bell and Linn's study showed to be effective in supporting students to use evidence in individual arguments, such as having students working in pairs and making thinking visible through reflective activities. In addition, it places great emphasis on dialogic argumentation for the development of argumentation skills. Previous research that utilized learning environments (LE) to support students' argumentation skills focused on scaffolding students through the LE to construct individual arguments. A distinctive feature of the present work is that students used the LE to engage in a series of dialogs with their peers and then were scaffolded by being prompted to reflect on the dialogs they had produced. Therefore in contrast to previous work which offered scaffolding for constructing an argument, in the present study students first produced their arguments, in the context of an authentic dialogic activity, and then were asked to reflect on and revise the arguments they had produced.

Another distinctive feature of the present study is the development of a web-based LE that included a rich content knowledge base regarding the intervention topic. In contrast to other learning environments that required students to review evidence from the World Wide Web (Bell & Linn, 2000), in the present LE students reviewed evidence from the knowledge base hosted inside the LE. The information included in the knowledge base had been adapted to be suitable to the cognitive level of the participants. Therefore, besides saving time for students, compared to surfing on the web to find relevant information, the knowledge base had the additional advantage of precluding the possibility that students did not comprehend the information they encountered in case that they would not use evidence in their arguments. The data presented in the LE were organized by thematic areas, listed in the Method section, not by the position they supported, because we wished to engage students in an authentic scientific activity. Engaging in the process of interpreting data to decide which position they might support, but also offering alternative interpretations to specific data presented by the opponent, in order to critique others' arguments, are important skills that students need to develop in order to engage in scientific argumentation.

Participants' argumentation skills were assessed before, during and after the intervention, to address our first research question: How does students' ability to use evidence change over time when they engage in an evidence-focused intervention based on argumentive dialogic activities and in some reflective activities? In addition, while views and findings regarding the role of adequate content knowledge for the development of argumentation skills are mixed (Hogan & Maglienti, 2001; Koslowski, 1996; Lawson, 2003; Sadler & Zeidler, 2005), in the present study we included a comparison group which had access to the same information—through access to the data base hosted in the LE—for the same amount of time as the experimental group but did not participate in dialogic argumentive activities. Our second research question was the following: Do students who engage in an evidence-focused dialogic intervention in a web-based learning environment exhibit any advantages regarding evidence-based argumentation when compared to students who do not engage in an argumentive discourse activity? The research design was a straightforward one in which participants were randomly assigned to one of two conditions, the evidence-focused dialogic condition or the comparison condition. Participants in the evidence-focused dialogic condition, working in pairs, engaged in a succession of dialogs with another pair who held an opposing view on the topic and in some reflective activities about the use of evidence in their dialogs. Participants in the comparison condition, also working in pairs, were asked to study the information in the learning environment's knowledge base and prepare a poster on the same topic that experimental-condition students worked on. Both conditions worked on the topic of Climate Change, one of the topics that science classrooms should discuss throughout K-12 education according to The Next Generation Science Standards (2013). Participants' argumentation skills were also assessed on another socio-scientific topic (Possible Fuels for Generating Electricity) to examine whether possible improvements observed in the intervention topic would transfer to a non-intervention topic.

Our hypothesis is that through dialogic argumentation students would come to appreciate the need for using evidence in their argumentation and they would increase their use of evidence. We coded students' dialogs in terms of evidence use and the function of evidence employed. Based on Walton's criteria of skilled argumentation and the findings of prior work focusing on argumentation strategies, which showed that sustained engagement in dialogic argumentation resulted in a shift in students' strategies from focusing on exposing their own position to critiquing the opponents' position (Iordanou, 2010), we coded evidence use to examine whether there was a change in the function of evidence use across time. In particular, we examined whether data were used to support one's own position or to weaken the opposing position, hypothesizing that there would be a shift in the function of evidence used from supporting own position to challenging opponents' position. In addition, given that an important objective of science education is to promote an understanding of how we have come to know what we know (Duschl, 2008) and the view that sustained argumentation can facilitate the development of epistemological understanding (Sandoval, 2005), we coded our data for evidence of gains in students' epistemological understanding. In particular, we examined whether students employed opinions or scientific data to back up their claims and whether they made explicit reference to data, by coding whether there was an explicit or an implicit reference to data and the source. Reference to the source of knowledge in argumentation is an implicit indication of an understanding that the source of knowledge matters when judging the trustworthiness of knowledge. We hypothesized that a shift would be exhibited from opinion to scientific knowledge and from implicit reference to data to explicit reference to data, including citation of the data's source.

Method

Participants

Thirty-two students (17 boys and 15 girls) took part in the study. Students were randomly assigned to the experimental and comparison conditions (16 students participated in each condition). Participants were students who volunteered to participate in a summer school organized by a public university in Cyprus. All were eleventh-grade high school students, 15- or 16-year-olds, from private and public schools of an urban area. Students were primarily from a middle-class population.

Initial and Final Assessments

Students' argumentation skills were assessed at both initial and final assessments through electronic dialogs. Two socio-scientific topics, climate change (intervention topic) and potential fuels for generating electricity (transfer topic), were used for assessment.

In order to assign students to sides for conducting the electronic dialogs, during initial assessment students' positions and supporting arguments regarding the two topics—Climate Change (CC) and Fuels for Generating Electricity (FGE)—were assessed individually via a written task, following a short passage introducing the scenario. The scenario regarding climate change was based on the scenario developed by Sadler, Chambers and Zeidler (2004). A new scenario was developed regarding the fuels for generating electricity (see Appendix A). Each scenario presented two opposing positions regarding the topic, human factors vs. natural factors for CC, and natural gas vs. coal for FGE. In addition, a short list of relevant facts was also administered to the students along with the two scenarios. An example of the facts provided for the FGE topic is "The production of 1 kilowatt of electricity releases 452 gr of carbon dioxide emissions into the atmosphere." No particular instructions were provided to students for using this information in their dialogs. The reason for providing this constrained knowledge base was to make sure that students had some data available if they wished to use them and also to exclude the possibility of lack of topic knowledge in case that they wouldn't use any data in their dialogs. Students were asked to indicate their position by choosing among three options: the two alternative positions of each topic and the option "Undecided" and to provide reasons to justify their choice.

Based on the participants' position on each topic, assessed in the earlier individual assessment, participants were assigned to sides for each topic. Participants who indicated that they were undecided gave reasons on both sides of the issue and were assigned to one or the other position in a way that served to equate the number of these participants on each side. Then participants had a dialog with a classmate who held an opposing position, for each topic. For the climate change topic, a student who endorsed the view that the causes of climate change were natural had a dialog with a student who endorsed the view that the causes of climate change were produced by humanity. For the electricity generation topic, a student who was in favor of coal as a source of generating electricity had a dialog with another student who was in favor of natural gas. Each student was assigned the same classmate to have a dialog with, on a particular topic, at both initial and final assessments. A different classmate was assigned to a particular student for each topic.

Dialogs were conducted via the Stochasmos chat tool (Kyza & Constantinou, 2007). Stochasmos is a web-based teaching and learning platform that can be used to design web-based learning environments for reflective inquiry. Students were instructed to conduct a dialog with the goal to persuade their interlocutors, who held an opposing position, that their own position was right and also to try to reach an agreement if they could. Dialogs lasted 15 minutes or less, if participants indicated they had finished. The software automatically saved the dialog for later analysis.

Intervention

Students were randomly assigned to one of two conditions, the evidence-focused dialogic condition—experimental condition—and the comparison condition. Students in both conditions worked on the same learning environment on the topic of Climate Change. The only difference between the two conditions concerned the activities in which students were involved. Each intervention took place during nineteen one-hour sessions occurring twice per day, four times a week, in three consecutive weeks. The intervention took place in a computer lab of a medium size public university within the context of a summer school program organized by the university.

Students in both conditions had access to a learning environment which was developed for the purposes of the present study. Each student, in both conditions, was paired with a same-side classmate, to work together throughout the intervention. The mission of students in both conditions was to get prepared for a conference, which would be conducted at the end of the intervention, to inform students and their parents about the causes of Climate Change. In particular, students in the experimental condition were instructed to prepare for a public debate regarding the causes of CC that would be conducted in the context of the conference, while students in the comparison condition were instructed to work in pairs to prepare a poster and a short oral presentation on the same topic of CC, which would also be presented at the same conference.

After the completion of the intervention, a conference was indeed organized, where all students who participated in the summer school and their parents were invited, in which students presented their work. Experimental group students presented a twenty minute, face-to-face debate, which was conducted at the group level, regarding the main cause of Climate Change, and attendees—students and parents—were asked to vote, by raising their hands after the completion of the debate, on which position was more convincing. Comparison-condition students presented their posters, along with short presentations, in a session that was organized at the same meeting.

The *Socrates* Web-based Learning Environment (LE)

A web-based learning environment was developed about Climate Change—the intervention topic—for the purposes of the present study. The LE was called *Socrates* after the ancient Greek philosopher, who stressed the importance of self-reflection for individual development. The LE was designed by a group of researchers and teachers working together. In particular, the group consisted of three researchers (two who majored in Science Education and one who majored in Educational Psychology) and four teachers (two elementary school teachers and two high school Biology teachers). Three of the teachers had graduate degrees in Science Education and one had a graduate degree in Educational Technology. Teachers had been selected based on their education, experience, and the interest they expressed following an open call. Teachers and researchers worked together as one group for the development of the learning environment. They had weekly meetings, over the course of six months, where they discussed the materials that would be included in the learning environment, the presentation of this information, and the cover story that would be used. Researchers had the leading role in the development of the educational curriculum, while teachers contributed substantially to the development of the knowledge base—finding relevant data and adapting them to be appropriate for high school students.

The *Socrates* learning environment was hosted in the platform of Stochasmos (Kyza & Constantinou, 2007). Stochasmos offers two main environments. The first is the Inquiry Environment, where a knowledge base for the topic of climate change was developed. The knowledge base included different types of information—short texts, graphs, tables and images (e.g., a graph of Earth's temperature over years). The second is the WorkSpace environment, which hosted the reflective templates "Finding Evidence," "Evidence for own argument," and "Evidence against other argument" (see intervention section below), where students were asked to construct evidence-based arguments and reflect on the arguments they produced while they were engaging in dialogic argumentation. Stochasmos offered students the opportunity to transfer information from the Inquiry Environment to the WorkSpace environment, and vice versa, using the “Data capture tool.” The platform also incorporated a chat tool, which was used for conducting students’ dialogic argumentation.

The *Socrates* learning environment consisted of five units, hosted under five different tabs (see Figure 1). The first unit, “Your Role” was an introduction to the learning activity and students’ mission. During the second unit, called “Greenhouse,” students studied global temperature change and the mechanism of the Greenhouse effect. This unit also included some experimentation for promoting further understanding of the Greenhouse effect. The third unit, called “Extreme Phenomena?”, presented data regarding potentially dangerous phenomena organized under the following categories (sub-tabs): Rainfall, Ice Formation , Sea level, Ocean temperatures, Desertification and Biodiversity. The fourth unit, called “Causes of Climate Change,” included data that could be used to hypothesize on the causes of Climate Change. An effort was made to have approximately equal sets of data supporting each position (human factors vs. natural factors). This unit included the following thematic areas (sub-tabs): Astronomical Phenomena, Earth, Oceans, Volcanoes, Paleoclimatology, Models and Tectonic Plates. The data presented here were of different formats, including graphs, figures, tables and short texts. Each piece of evidence included information about its source. Examples of data presented in this unit were a graph showing the atmospheric CO_2 levels from 1960 to 2010 from Wikipedia and a figure showing the surface temperatures over the last 1,100 years from the National Academy of Sciences website. After each table, figure, or graph a short text followed which briefly descripted the data. The inclusion of this short description aimed to help students overcome any difficulty they might have had in interpreting data in different formats and therefore rule out the possibility that participants had difficulty in interpreting data in the event that they would not use data adequately.

The final unit was about the socio-economic aspects of Climate Change. The following sections were included in the fifth unit: Economy, Intergovernmental Panel on Climate Change (IPCC), Research, the Kyoto Protocol and opinions about it, and Politics. An example of a piece of evidence included in this section is the following “Environment ministers preparing for next week’s talks on global warming in Poznan, Poland, have been sounding decidedly downbeat. From Paris to Beijing, the refrain is the same: This is no time to pursue ambitious plans to stop global warming. We can’t deal with a financial crisis and reduce emissions at the same time” (Source: New York Times, 27.11.2008).

An initial version of the *Socrates* learning environment was pilot tested in a sample of eleventh-grade, 15–16 year-old, high school students in a public school in the same country as the one in which the study intervention took place. After the pilot implementation, we revised the learning environment based on students’ and teachers’ feedback, as well as our own observations. The revisions concerned mainly the content of the knowledge base: we reduced the information presented, we shortened lengthy texts, and we added explanations by creating an electronic glossary in the learning environment for terms that students might not have been familiar with.

Figure 1.

Procedures

The first four sessions were identical for the two conditions. During these sessions students were introduced to the problem by studying recent extreme weather phenomena and were asked to prepare a short hand-out describing three of them. In addition, students examined the mechanism of the Greenhouse effect. Although students had examined the Greenhouse effect mechanism as part of their school curriculum in their seventh grade, we decided to provide this information to refresh their memory about the phenomenon since this information would be important for their conceptual understanding of the information that they would study later in the intervention. From the fifth session on, students in the intervention condition engaged in different activities from those of the comparison-condition students.

Experimental Condition.

Finding Evidence. In the fifth, sixth and seventh sessions (see Figure 2), experimental-condition students were asked to review the information included in unit four of their LE and construct evidence-based arguments, with the help of the "Finding Evidence" reflection sheet. The purpose of preparing those arguments, they were told, was to prepare for a series of discussions that would follow. The "Finding Evidence" reflection sheet asked students to state a claim and to provide evidence from the LE to support their claim. A separate reflection sheet was used for each argument they made. All the reflection sheets constructed were saved by the system in each student's account to be available for students to access when they would engage in electronic discussions. The "Finding Evidence" reflection sheet included a picture of a tree which researchers used to show the function of evidence in the context of an argument. The role of evidence in the context of an argument was compared with the role of roots in a tree. This activity was repeated during the 12^{th} and 13^{th} sessions (see Figure 2), where students were asked to review unit 5, which included information regarding the politico-economic aspect of Climate Change. During this activity students used the Stochasmos Data Capture Tool, which enabled them to capture segments of the data and automatically transport them to the WorkSpace area where they could use them in preparing their arguments.

Argumentation. The major activity of the experimental-condition curriculum was dialogic argumentation. Students who participated in the experimental condition, working in pairs, engaged in a series of electronic dialogs with an opposing side pair during the 7th to 11th and 12th to 15th sessions (see Figure 2). In each of these sessions, participants had discussions with a different opposing pair. Participants were asked to engage in these dialogs with the goal to persuade their interlocutors, who held an opposing position, that their own position was right. After the completion of each dialog, participants reflected on an electronic transcript of their dialog, with the help of two electronic reflection sheets the "Evidence for own Argument" and the "Evidence against other Argument." These reflection sheets were based on reflection sheets that were used in prior studies (Iordanou, 2010; Kuhn, Goh, et al., 2008). However, in addition to asking students to reflect on whether they had addressed the opposing side's claims, the reflection sheets used in the present study prompted students to also reflect on whether they had used evidence to support their critique. In particular, the "Evidence against other argument" reflection sheet asked students to reflect on the effectiveness of the counterarguments they produced and the evidence they used to support their counterarguments. The "Evidence for own argument" reflection sheet prompted students to reflect on the Rebuttals they offered and the evidence they used, to counter the opposing pair's counterarguments and increase the strength of their own argument. Before the completion of each reflection sheet the terms "counterargument," defined as the opposing side's argument against their own argument, and "rebuttal," defined as their response (counterargument) to the counterargument offered by the opposing side to their own argument, were introduced to students. When pairs completed their templates, they exchanged templates with another pair who held the same position for providing and receiving feedback.

Preparation and Showdown. In the 16th and 17th sessions, students working in groups of four, reviewed and reflected on the reflection sheets they had completed in previous sessions in order to prepare evidence-based arguments that they could use to support their position and the critiques they offered to the opposing position. Each group was working with a research assistant, who facilitated the discussion. Printed transcripts of their reflection sheets and colored index cards were made available. Students were encouraged to use a particular colored card for evidence and make sure that each argument they produced included cards of that particular color.

The culminating point of the curriculum activities was a class-wide electronic debate between students holding opposing positions. Students supporting opposing positions were located in different rooms and communicated through the Stochasmos platform. The dialog was projected on a wall screen so all students had the opportunity to participate. In the session following the Showdown, an argument map was presented to the students, prepared by the researchers, where different colors were used to indicate effective and less effective argumentive moves as well as whether evidence was used effectively to support students' arguments and critique.

Comparison Condition.

Participants in the comparison condition were asked to study units four and five of the *Socrates* LE and, working in pairs, to prepare a hard-copy poster or a model describing the causes of climate change. Research assistants were available to offer help, when needed, regarding conceptual clarifications or other problems regarding technical issues that students might have had regarding the Stochasmos software. When pairs completed their project, they shared their work with another pair in the comparison condition to provide and receive feedback.

Figure 2.

Coding

The dialogs of both groups were segmented into idea units, with an idea unit defined as an assertion with any possible accompanying justification that might have been provided. Then, idea units were coded as to whether or not they included evidence. A statement was considered as evidence if it offered an implicit answer to the question “how do you know,” and if the answer came from an available source (rather than someone's mind as in an opinion or an assumed fact). Each idea unit containing evidence was coded based on its function and accuracy. Using an adapted version of Kuhn and Moore’s coding scheme (in press), each unit containing evidence was coded into one of three categories, based on its function: a) evidence employed to support a claim, b) evidence employed to weaken a claim and c) meta-level talk about evidence (see Table 1). Evidence employed to weaken a claim was further coded as “Weakenopp_C”, if the evidence used critiqued and removed power from the opponents’ claims, or “Weakenopp_A” if the evidence used served as an alternative argument. A further distinction that emerged from students’ work was a distinction between idea units containing quantitative evidence and idea units containing non-quantitative evidence. We thought that this was an important distinction, since employment of quantitative evidence would make a scientific argument more convincing than employment of non-quantitative evidence. Idea units including numerical data were coded as quantitative evidence whereas idea units that did not include any numerical data were coded as non-quantitative evidence. For example, “Coal’s value corresponds to 1/10 of the price of other raw materials” was coded as quantitative evidence, whereas “coal is less expensive” was coded as non-quantitative evidence. Each unit was also coded based on its accuracy in one of the following categories (a) undocumented evidence claims from personal knowledge, (b) distorted use of data from the LE, when students misinterpreted or misused data from the LE, and (c) correct use of data from the LE or from personal knowledge, when students used appropriate data from the LE to support their claims. The third category of correct use of data was further differentiated into explicit reference to data from the LE and their source, when participants cited data from the LE and made a specific reference to the source of the data (e.g., "as you can see in graph A1"), and into implicit reference to data from the LE, when participants used information from the LE without mentioning the source of the information. Two coders, blind to the treatment, time, and identity of the participants, participated in segmenting and coding. Reliability was calculated, based on 25% of the data, with Cohen's kappa. Kappa was 0.86 for the segmentation of idea units and 0.9 for coding evidence function, indicating good reliability.

Table 1.

Results

To examine our first research question regarding how students' use of evidence changed over time when engaged in an evidence-focused dialogic intervention, we first examined experimental-condition students' argumentation skills before and after their engagement in the intervention. Then we examined the experimental-condition students' argumentation skills over time during the intervention to examine possible connections between specific design features employed in the intervention and gains in argumentation skills. Finally, a comparison between experimental- and comparison-condition students was conducted to address our second research question regarding whether students who engaged in an evidence-focused dialogic intervention would exhibit any advantages in evidence-based argumentation compared to students who had access to the same information, for the same amount of time, as the experimental group but had not participated in dialogic argumentive activities.

Use of Evidence at Initial and Final Assessment

Students' use of evidence in argumentation was examined in the intervention and transfer topics. For each topic, we examined use of evidence at the initial and final assessments focusing on the percentage of idea units that included evidence, including an examination of the percentage of idea units that included empirical evidence, and the function of each piece of evidence employed. The accuracy of evidence employed was also examined in a separate analysis. The unit of analysis was the individual student participant. Because there was a difference in the number of utterances produced at the initial and final assessments, percentages of usage were calculated for each participant, rather than frequencies. An arcsine square root transformation, which normalizes proportional data (Cardinal & Aitken, 2013), was performed to normalize these proportions. Three students were absent during the final assessment of the intervention topic and therefore were excluded from the analysis of the intervention topic.

Intervention Topic

Evidence Use. The overall use of evidence, as well as the function of evidence used, in argumentation of experimental-condition students are summarized in Table 2. We used paired-sample t tests on each coding category to estimate effects of the intervention on students' ability to use evidence in argumentation. As seen in Table 2, experimental-condition students doubled the usage of evidence from initial to final assessment, $t(11) = 2.35$, $p = .039$, *Cohen's d* = .68. A significant increase was also observed in the idea units containing quantitative evidence, $t(11) = 9.03$, $p < .001$, *Cohen's d* = 2.61. Participants increased the percentage of idea units containing quantitative evidence from 4.64% (SD = 5.06) at initial assessment to 44.99% (SD = 9.49) at the final assessment.

Function of Evidence. Paired-sample t tests on the use of the overall weakenopp category, including both the Weakenopp_C and the Weakenopp_A categories, and on

Weakenopp_C category in particular showed that students made significant gains in using evidence to critique the opponents' position, $t(11) = 5.536$, $p < .001$, *Cohen's d* = 1,60, and $t(11) = 6.47$, $p < .001$, *Cohen's d*=1,87, respectively. As seen in Table 2, students exhibited an increase in the percentage of idea units which functioned to weaken others' claims, from 5.31% (*SD* = 6.40) to 23.33% (*SD* = 7.55), and in particular, in the idea units which directly critiqued the opposing position (Weakenopp_C), from 3.21% (*SD* = 4.71) to 18.75% (*SD* = 7.79). Note that the results of the experimental condition were also confirmed by repeated measures ANOVA (see section on Experimental vs. Control condition students' argumentation skill on the intervention and transfer topics, below), which eliminates the possibility of a cumulative Type I error due to multiple testing.

In terms of the percentage of idea units containing evidence which functioned to support one's own claim, no significant change was observed from initial to final assessment. As seen in Table 2, using evidence to support one's own position was the most common function of employing evidence at initial assessment, 15.56% (*SD* = 8.08), and remained a popular practice at the final assessment, 20.25% (*SD* = 11.48). The percentage of idea units containing meta-level talk about evidence was very limited, less than 5%, at both initial and final assessments and no significant change was observed from initial to final assessment.

Table 2.

Percentage of evidence that functioned to support own position and weaken opponent's position.

To examine further how evidence were used in dialogs we focused only on the idea units containing evidence and we compared the percentage of evidence which functioned to support one's own position and the percentage of evidence which functioned to critique the opposing position at initial and final assessment. At initial assessment, of the evidence units used, the majority, 70%, functioned to support one's own position and only 30% of the evidence units served to weaken opponents' positions. At the final assessment, the percentage of evidence units which functioned to critique the opposing position increased from 30%, at initial assessment, to 51%, at the final assessment ($p <.001$, McNemar). The percentage of idea units functioned to support one's own position decreased from 70% at initial assessment to 45% at the final assessment ($p <.001$, McNemar). These findings showed a shift in participants' attention from one's own position to the opponent's position.

Accuracy C-Codes

As seen in Table 3, the majority of evidence units produced at initial assessment, 79%, came from general knowledge. Only a small percentage, 10%, of the evidence units produced, included accurate data from the learning environment. By the end of the intervention, though, students increased significantly the correct use of data from the LE, from 10% at initial assessment to 59% at the final assessment ($p <.001$ McNemar test). A significant increase was also observed in students' performance regarding the percentage of idea units in which there was an explicit reference to evidence from the LE, from 2.6% at initial assessment to 53% at the final assessment ($p < .001$ McNemar test). While experimental-condition students increased the percentage of idea units that made correct use of data from the LE, they decreased the percentage of idea units that included undocumented evidence claims from personal knowledge, from 79% at initial assessment to 29% at the final assessment ($p =.001$, McNemar test).

Table 3.

Evidence Use Over Time During the Intervention

To examine possible connections between specific design features employed in the intervention and gains in argumentation skills, we examined the participants' use of evidence in argumentation over the course of dialogs taking place during the intervention. In particular, the participants' use of evidence in argumentation was examined after participants completed the finding evidence scaffolding activity (Chat 1, session 7, see Figure 2), after they engaged in two electronic dialogs with an opposing side pair (Chat 2, session 9), after they completed the first reflective activity where, with the help of the "Evidence for own argument" reflection sheet, students were asked to reflect on the effectiveness of the counterarguments they produced and the evidence they used to support their counterarguments (Chat 3, session 10), and after they completed the second reflective activity, where students with the help of the "Evidence against other argument" reflection sheet were prompted to reflect on the Rebuttals they offered (and the evidence they used) to counter the opposing pair's counterarguments (Chat 4, session 11). Table 4 presents the percentage of idea units containing evidence and the function of evidence used across time. We compared participants' performance on each coding category between chat 1 and chat 2, chat 2 and chat 3 and chat 3 and chat 4, to identify possible differences in participants' use of evidence in argumentation over time. Note that the unit of analysis for the dialogs conducted during the intervention was the pair, since participants were working in pairs. Because of the small number of pairs (N = 8) a non-parametric test was used to examine participants' argumentation skills over the course of the intervention. As seen in Table 4 there was a significant increase in the idea units containing evidence from chat 2, 27.09% (SD = 5.01) to chat 3, 57.33% (SD = 16.23), Z = -2.380, Wilcoxon test, p =.017, effect size r = .842. In addition, there was a significant increase from chat 2 to chat 3 in participants' percentage of idea units which critiqued the opponent's position, from 10.35% (SD = 7.18) in Chat 2 to 27.60% (SD

= 8.55) in Chat 3 (Z = -2.521, p = .012, effect size r = -.891) and in the percentage of idea units which contained the advanced weakenopp_C strategy, from 7.99% (SD = 6.14) to 23.91% (SD = 7.13), (Z = -2,521, p =.012, effect size r = -.891). In addition to gains at the procedural level, regarding the function of evidence used, gains were also observed at the meta-level. Participants' idea units which contained meta-level talk about evidence increased from 3.24% (SD = 1.77) in Chat 2 to 11.82% (SD = 7.18) in Chat 3 (Z = -2.100, p = .036, effect size r = -.707). Participants' gains in argumentation from Chat 2 to Chat 3 suggest that engagement in a reflective activity about the use of evidence in argumentation, which was the new element that was introduced after Chat 2, might have had a positive impact on participants' use of evidence in argumentation. We return to this finding in the discussion section.

Table 4.

Transfer Topic

We used paired-sample t tests on each coding category to examine whether experimental-condition participants transferred their gains in evidence use in argumentation observed in the intervention topic to the transfer topic.

Evidence Use. Regarding overall use of evidence in argumentation, a significant increase was observed from initial to final assessment, $t(11) = 2.646$, p =.023, *Cohen's d* = 0.764. In particular, experimental-condition students increased the percentage of idea units containing evidence, as seen in Table 5, from 26.75% (SD = 11.08) at initial assessment to 39.83% (SD = 13.12), at the final assessment.

Function of Evidence. Regarding the function of evidence, paired-sample t tests showed that there was a significance increase in the use of the Weakenopp_C category from initial to final assessment, $t(11) = 3.782$, $p = .003$, *Cohen's d* = 1.092. As seen in Table 5, the percentage of idea units that were coded as Weakenopp_C at initial assessment was 2.31% (*SD* = 5.53), whereas the corresponding percentage at the final assessment was 13.48% (*SD* = 7.31). No significant change was observed from initial to final assessment regarding use of quantitative evidence, Overall Weaken-opp and Support-own categories.

Table 5.

Individual Performance

In addition to analyses of group trends, equally informative are analyses of changes at the individual level. Table 6 shows the percentage of participants who produced at least one idea unit containing evidence, containing evidence which functions to weaken opponent claims—including both the overall_weakenopp category and the more accomplished weakenopp-C category—and containing evidence which functioned to support own claim, at initial and final assessment, in intervention and transfer topics. As seen in Table 6, only 17%—2 out of 12—of experimental condition participants produced idea units containing evidence which functioned to weaken opponent's claims by critiquing them (weakenopp_C), at both the intervention and the transfer topic, whereas at the final assessment, all—12 out of 12—and almost all of them—11 out of 12—did so at the intervention and transfer topic, respectively. In addition, a significantly greater percentage of participants produced idea units containing evidence at the final assessment—100%—compared with initial assessment—50% (p =.031, McNemar)—as well as idea units containing evidence which functioned to weaken opponents' claims, including both weakenopp_A and weakenopp_C categories, from 17% to 100% (p =.002, McNemar Test) on the intervention topic.

Table 6.

Accuracy of evidence codes used.

At initial assessment, the majority of evidence units produced by experimental-condition students included undocumented evidence from students' personal knowledge, as seen in Table 3, while only a small percentage of the evidence units produced made explicit reference to accurate data from the LE.

At the final assessment, though, participants increased significantly the percentage of evidence units which made explicit reference to data from the LE, from 6.50% to 21.10% (p =.001, McNemar).

Experimental vs. Control condition students' argumentation skill on the intervention and transfer topics

In order to examine our second research question, whether students who engaged in an evidence-focused dialogic intervention would exhibit any advantages in evidence-based argumentation compared to students who had access to the same information for the same amount of time as the experimental group but did not participate in dialogic argumentive activities, we conducted a comparison between experimental and control condition students. To test the effect of conditions, a 2 X 2 (Condition X Time) repeated-measures ANOVA was performed, with Time as the within-subjects variable and Condition as the between subjects variable, for each coding category in both the intervention and transfer topics.

Intervention Topic.

Experimental group participants contributed an average of 16.08 (SD = 5.98) idea units at initial assessment and 13.42 (SD = 3.99) units at the final assessment. Comparison group participants contributed an average of 14.00 (SD = 7.63) idea units at initial assessment and 17.77 (SD = 13.08) idea units at the final assessment.

Evidence Use. A 2 X 2 (Time X Condition) repeated-measures ANOVA for the idea units containing evidence revealed a two-way interaction $F(1, 23) = 12.456$, $p < .001$, partial $\eta^2 = .351$. As seen in Table 2, experimental group participants used more evidence at the final assessment than at initial assessment, whereas no change was observed in comparison group's performance, from initial 24.86% (SD = 16.58), to final assessment, 22.23% (SD = 14.35). A 2 X 2 (Time X Condition) repeated-measures ANOVA for the idea units containing quantitative evidence also revealed a two-way interaction, $F(1, 23) = 29.38$, $p < .001$, *partial* $\eta^2 = .561$. While there was an increase in experimental condition's idea units containing evidence, from 4.64% (SD = 5.06) at initial assessment to 44.99% (SD = 9.49) at the final assessment, no change was observed in comparison-condition students' idea units containing evidence from initial, 5.79% (SD = 6.06) to final assessment 7.52% (SD = 7.53).

Evidence Function. An analysis of use of overall weakenopp category, including both Weakenopp_C and Weakenopp_A, showed a significant Time X Condition interaction, $F(1, 23) = 7.36$, $p = .012$, *partial* $\eta^2 = .242$. Experimental-condition students, as seen in Table 2, exhibited an increase in the percentage of idea units which functioned to weaken other's claims, from 5.31% (SD = 6.40) to 23.33% (SD = 7.55), whereas comparison-condition students showed no change in the respective percentages from initial—6.10% (SD = 9.41)—to final assessment—10.08% (SD = 11.68). A two-way ANOVA analysis on Weakenopp_C category, in particular, showed a Time X Condition interaction, $F(1, 23) = 12.456$, $p < .001$, *partial* $\eta^2 = .351$. As in the analysis of overall weakenopp category, only participants in the experimental condition showed an increase in Weakenopp_C usage, from 3.21% (SD = 4.71) to 18.75% (SD = 7.79), whereas comparison condition participants showed no significant change, from 1.23% (*SD* = 3.01) to 6% (*SD* = 8.49). No significant difference was observed between experimental and control condition students regarding the Support-own category and the meta-level talk about evidence. In particular, the percentage of idea units containing evidence that functioned to support one's own claim was high at both initial,16% (8.08) and 12% (10.48), and final assessment, 20% (*SD* = 11.48) and 10% (*SD* = 9.30), in both experimental and control conditions, respectively. The percentage of idea units containing meta-talk about evidence was very limited, less than 5%, at both initial and final assessment in both conditions, and no significant change was observed from initial to final assessment.

Accuracy-C Codes. Comparison-condition students, like the experimental-condition students (see Table 3), increased the percentage of idea units that made correct use of data from the LE, from 26.5% to 43.3% (p <.001 McNemar test). Yet, in contrast to the experimental-condition students who exhibited an increase in the percentage of idea units in which there was an explicit reference to evidence from the LE, comparison-condition students increased the percentage of idea units that included an implicit reference to evidence from the LE (p <.001 McNemar test). No change was observed in the percentage of comparison condition's idea units included undocumented evidence claims from personal knowledge, which was high at both initial—70.6%—and final assessment—45.3%.

Transfer Topic

Experimental group participants contributed an average of 15.08 (*SD* = 7.28) idea units at initial assessment and 15.17 (*SD* = 5.42) units at the final assessment. Comparison group participants contributed an average of 8.94 (*SD* = 4.39) idea units at initial assessment and 12.44 (*SD* = 6.26) units at the final assessment.

Evidence Use. A two-way ANOVA on the percentage of idea units containing quantitative evidence revealed a significant Time X Condition interaction, $F(1, 26) = 14.768$, $p = .001$, *partial* $\eta^2 = .362$, while a 2 X 2 (Time X Condition) repeated-measures ANOVA on overall evidence use showed no significant Time X Condition interaction. The percentage of idea units containing quantitative evidence at initial assessment was very small, less than 5%, in both the comparison and experimental group participants' dialogues. At the final assessment, however, experimental group participants increased their usage of quantitative evidence, while comparison condition participants did not exhibit any change as compared to initial assessment. In particular, the percentage of idea units containing quantitative evidence in comparison-condition participants was 3.23% (*SD* = 6.27) at initial assessment and remained about the same at the final assessment, 3.00% (*SD* = 4.87). Similarly, no change was observed in control-condition students' percentage of idea units containing overall evidence from initial, 30.56% (*SD* = 15.06) to final assessments, 31.12% (*SD* = 15.98).

Evidence Function. A two-way ANOVA analysis on each of the categories of Overall Weakenopp, Weakenopp_C and Supportown showed a significant Time X Condition interaction for the Weakeneopp_C category, $F(1, 26) = 6.32$, $p = .018$, *partial* $\eta^2 = .196$. While experimental-condition students, as seen in Table 5, exhibited an increase in the percentage of idea units that functioned to weaken other's claims, from 2.31% (*SD* = 5.53) to 13.48% (*SD* = 7.31), comparison-condition students showed no change in the respective percentages from initial—3.50% (*SD* = 7.79)—to final assessment—5.69% (*SD* = 8.43). No significant Time X Condition interaction was observed regarding the Overall Weakenopp and Supportown categories. Control condition’s percentages of usage of the Supportown category and the Overall Weakenopp Category at initial, 20.10% (*SD* =11.94) and 10.45% (*SD* = 9.62), and final assessment, 18.28% (*SD* = 12.89) and 12.84% (*SD* = 11.46) respectively, were comparable with the corresponding percentages of the experimental-condition students (see Table 5).

Accuracy-C Codes. No significant change was observed in the comparison-condition students’ performance from initial to final assessment in terms of accuracy of evidence used. 41.8% and 46% of control condition participants’ idea units contained correct use of data from the LE at initial and final assessments. The percentage of control condition participants’ idea units that contained undocumented evidence claims from personal knowledge was 44% at initial assessment and remained high at the final assessment, 40%. Unlike the experimental-condition students who exhibited a significant change in explicit reference to evidence (see Table 4), no change was observed from initial to final assessment in control condition’s explicit reference to data, from 6.9% to 11%, or implicit reference to data, from 24.9% to 35%.

Examples of Students’ use of evidence in argumentation

In this section, we provide illustrative examples of participants’ dialogs to exemplify the changes observed in evidence use. Students are identified using pseudonyms. Examples from dialogs on Climate Change, conducted at initial assessment and final assessment by experimental-condition students are presented and discussed.

Example of students’ dialog on Climate Change at the Initial Assessment

Nick I support that it is a natural phenomenon

Chris I support that it is human responsibility

Nick Why?

Chris Well, I think that if there weren't any human actions the Earth's temperature would not increase

Nick I believe that it is a natural phenomenon.

Nick Because it happened in the past when humans did not release that many natural gases.

Nick It is not only human responsibility… the Earth is also responsible. Humans are not responsible for everything

Chris I know that humans are not responsible for everything. I mean with the way they act and they take advantage of the gases they may contribute (to the problem)

Nick Why are they responsible for this? I don't understand what you mean

Chris I mean that humans are not responsible for everything, but they are responsible for this one, because if we examine the most recent studies we will see that with the progress of technology and the invention of how to use gases, the number of gases increased substantially as well as the Earth's temperature.

Chris So humans are responsible for the increase in temperature.

Nick Maybe you are right.

During the initial assessment, as seen in the dialog above, most evidence used served to support student's own position. Students appear to focus their attention on how to back up their own position in order to convince the opponent of the superiority of their position. Both students initially stated their positions and then provided evidence supporting their own positions, either after being asked by the opponent, in the case of Chris, or as a response to the opponent after he or she provided justification to support his or her own position, in the case of Nick. Note that although students expressed disagreement, as when Nick said “It is not only human responsibility… the Earth is also responsible. Humans are not responsible for everything,” they didn’t employ evidence to support their disagreement.

In terms of the accuracy of the evidence used at initial assessment, as seen in the example above, the evidence was either based on personal opinion (e.g., I think that if there weren't any human actions the Earth’s temperature would not increase) or general knowledge (e.g., it happened in the past). In the case where the evidence used made correct use of data from the LE, such as Chris’ last response, “So humans are responsible for the increase in temperature,” there was an implicit reference to data, eg, referencing the most recent studies without defining the studies a student was referring to.

Example of students’ dialog on Climate Change at the Final Assessment

Andrew	Natural causes
John	Human made causes!
Andrew	Why do you claim that humans are responsible for climate change?
Andrew	Let’s start with livestock. Methane is one of the major greenhouse gases which however can be found in only very small quantities, almost negligible
John	There are 55 million cattle in the world which are bred and humans are responsible for this increase. The Argentine researcher Guillermo Berra estimated that much of the greenhouse gas emissions in the country could be methane emitted by 55 million cattle grazing on the plains. Specifically, to examine how the cows release gas, he hang huge balloons on the backs of several cows,

according to Reuters. The researchers could never imagine, according to Reuters, that one cow can produce 28-35 cubic feet methane every day.

Andrew Cows produce methane indeed, but the percentage of methane is very small. In fact, steam represents the 2/3 of greenhouse gases, a gas which is pure natural.

At the final assessment, as seen in the dialog above, students' attention focused on using evidence to weaken the opponents' position. For example, when John made his point that there are 55 million cattle in the world releasing methane, Andrew used evidence that functioned to directly address John's argument (Weakenopp-C) by saying that the concentration of methane in the atmosphere is very small. That is, Andrew used evidence to directly critique John's argument and reduce its power. Andrew continued his critique, "in fact, steam represents the 2/3 of greenhouse gases, a gas which is pure natural," by offering evidence that served as an alternative argument to John's argument (Weakenopp_A).

Not only students used evidence to critique others' positions at the final assessment, but they also cited the source of their evidence, making their critique more convincing. The following is another example of students dialog at the final assessment.

Example of students' dialog on Climate Change at the Final Assessment

Andrew Sudden changes in temperature were observed since ancient times, even before humans developed the technology, built factories and took action, as they claimed, for the Earth's climate. We can see this in graphs from paleoclimatology. Further, in times when the concentration of CO_2 was very high, the temperature was lower than other times when the concentration (of CO_2) was lower. Therefore you can't prove that the gases have an effect on climate.

John According to graph 2 around 1800 there is a rise of CO_2 and at the same time a rise of temperature. There (at that time) the industrial revolution began and humans affected the environment with their actions. Even before 1000 years(,) according to History(,) humans made use of forests for smelting copper and creating currency

Andrew But if you see (in graph 2) between 1350 to 1450 the percentage of CO_2 was high and the temperature was at the same level

As seen in the example above, students made explicit reference to data from the LE, such as John's reference to graph 2 and Andrew's reference to graphs from paleoclimatology. Andrew employed evidence from graphs in paleoclimatology to show that sudden changes in the temperature had been reported before human industrial revolution and that there was no direct relation between high concentrations of CO_2 and rise in temperature, supporting his position that climate change is natural. John directly critiqued Andrew's argument by employing evidence from another graph (graph 2) showing that an increase in CO_2 and a rise in temperature were observed with the beginning of the industrial revolution, critiquing John's argument that there is no relation between concentrations of CO_2 in the atmosphere and temperature rise. Andrew, then, critiqued John's position by giving an alternative interpretation to the same graph (graph 2) that John used to support his own argument. In particular, by pointing out that in graph 2 there was a period when the concentration of CO_2 was high but the temperature was stable, Andrew provided evidence that functioned to directly critique John's argument in favor of a direct relation between CO_2 and temperature.

Discussion

The aim of this study was to examine how students used evidence in argumentation when they engaged in a series of argumentive and reflective activities in the context of a web-based learning environment. Our findings showed that 11th graders exhibited advancements in using evidence in argumentation, after they engaged in a computer-supported argumentive discourse activity. In particular, by the end of the intervention experimental-group students increased the use of evidence in their dialogs, used more evidence that functioned to weaken opponents' claims and used more accurate evidence, making also explicit reference to the source of the evidence, than they did at initial assessment. Notably, experimental-group participants transferred these findings to another, non-intervention, socio-scientific topic. Participants in the comparison group did not exhibit any improvements in either the overall usage of evidence or the function of evidence they employed from initial to final assessment. The improvements found in the present study, in contrast to findings from earlier research (Osborne et al., 2013; Erduran et al., 2004), show that students' argumentation skills are amenable to development when students receive specific attention in the context of a student-centered dialogic intervention. These findings are in line with previous research showing the merit of sustained engagement in argumentive dialogic activities (Asterhan &Schwarz, 2007; Kuhn, Shaw & Felton, 1997; McNeill, 2009; Ryu & Sandoval, 2012) in promoting students' argumentation skills. For example, Asterhan and Schwarz (2007) showed that engagement in argumentive discussion yielded greater learning gains regarding understanding of scientific concepts than engagement in mere collaborative activities.

What induced the observed improvements? Our analysis of participants' dialogs over the course of the intervention was particularly insightful in revealing considerable gains in evidence use in argumentation after participants engaged in reflective activities about their dialogs. This finding suggests that reflective activities on evidence use might be an important feature of the intervention that supported the development of students' ability to use overall more evidence in argumentation and particularly evidence that functioned to critique others' positions. This finding, suggesting a facilitative role of reflective activities in promoting evidence use in argumentation, is in line with previous research showing that reflective activities focused on employment of counterargument strategies supported the development of participants' counterargument and rebuttal strategies (Felton, 2004), and extends previous work by showing that reflective activities focused on evidence use in argumentation, along with engagement in argumentation, can support participants in employing evidence to back up their arguments and counterarguments. Note that in contrast to previous studies using technology-enhanced environments to support argumentation that offered scaffolding for constructing a written argument (Bell & Linn, 2000; Clark et al., 2008), in the present study "scaffolding" followed students' engagement in a dialog. Students, instead of being asked *to use* evidence in their arguments, were asked *to reflect on whether they had used* evidence in their arguments after they had been engaged in dialog with peers and *to revise* their arguments, in respect to employing evidence to back up their claims. Yet, we believe that for interpreting the gains observed in argumentation in the present study, we cannot focus on a single component of the intervention, ignoring the broader context in which this component was embedded. To gain a fuller understanding of what has contributed to the gains observed we should take into consideration that students engaged in a goal-based activity in the social context that we believe made the goal of persuasion in argumentation, and particularly the persuasive role of evidence in argumentation, meaningful.

How were these improvements achieved? The increase in meta-level discourse observed over the course of intervention dialogs, as well as the improvements observed in explicit reference to the data source from initial to final assessment, suggest that there were gains in meta-level understanding of argumentation, both meta-strategic and epistemological, besides the gains observed at the procedural level. We see the meta-level discourse as significant in indicating at least an implicit understanding of the objectives of argument and particularly of the purpose of using evidence in argument. When participants say to their interlocutors, ‘‘Give us some evidence’’ or ‘‘You have not provided evidence,’’ they are displaying implicit understanding of what contributors to a dialog need to do in order to engage effectively and convincingly in argumentation.

It is notable that these increases in meta-level discourse were evident only when participants were engaging in the dialogic task with a collaborating partner, during the intervention, not when participants worked without a partner (at the initial and final dialogic assessments). This finding is aligned with Kuhn, Goh, et al. (2008) findings of improvements in meta-level communication observed when participants worked in a collaborative context but not when they were working individually. Besides the cognitive and social support for one another when working in the collaborative context, students also shared the social goal of prevailing over the opposing side in the showdown. The absence of both these conditions when students argued individually makes plausible the interpretation that students perceived no need to express themselves in meta-level talk in the solitary condition. If so, our findings point to the importance of social context and goal structure in influencing scientific argumentation—a conclusion also reached by Berland and Reiser (2011) who found that students who argued with the goal of persuading engaged more in evaluation of others' arguments than students who argued with the goal of sense-making.

Students not only exhibited gains in meta-level discourse about the use of evidence in argumentation, they also exhibited gains in epistemological understanding. We see the increase in the use of scientific data to support students' arguments and the decrease in employment of personal opinions, as well as the increase observed in citation of the source of the data used, as indicators of students developing, at least implicitly, an epistemological understanding of the constructive nature of scientific knowledge. Students exhibited a shift from presenting their "right," self-evident theories of how things are, without providing any data to support their arguments beyond presenting their personal opinions, to employing data to support their positions and offering alternative interpretations for a particular piece of evidence. These findings have two important implications. The first implication is that dialogic argumentation can be a suitable setting for studying students' epistemological beliefs, particularly the process of change. Dialogic argumentation offers a window to students' epistemological beliefs and to what Chinn and colleagues (2011) refer to as the students' epistemic aims and values. Although it was not in the objectives of the present study to examine how epistemic beliefs change, an important implication of our findings is that studying students while engaging in dialogic argumentation over time could be a promising way to respond to the need, highlighted by Sandoval (2014), for more research employing the micro-genetic method in order to develop a fuller picture of epistemological development. The second implication is that engagement in argumentation appears to be a promising way to support the development of a constructivist epistemological understanding in science (Iordanou, 2010; Sandoval, 2005). Future research, including measures that directly assess students' perspectives on scientific knowledge and the process of knowing in science, is required to provide further evidence for the relation between argumentation and epistemological understanding.

Turning to the changes observed at the strategic level regarding using evidence in argumentation and the function of evidence used, it is notable that positive change over time in evidence use did not diminish when participants returned to the solitary condition of working without a collaborating partner and without a social goal. An important implication of this finding is that argumentation in the social plane can serve as a promising pathway for the development of scientific argumentation at the individual plane. Notably, participants in our study used evidence not only to support their own claims, but also to critique the opponents' claim, which Walton (1989) identifies as a central objective of skilled argumentation. This finding is in line with previous work showing that after engagement and practice in argumentive dialogic and reflective activities, students shifted their attention from their own position to the opponents' position (Kuhn, Goh, et al., 2008; Kuhn et al., 2013). Instead of trying to convince others of the superiority of their own positions, students tried to identify flaws in the opponents' positions, through evidence-based counterarguments.

Despite the improvement observed in the experimental condition in the use of evidence in argumentation, the percentage of idea units containing evidence remained below 50% even after students had participated in the dialogic intervention. Along with these findings are the findings regarding percentage of idea units containing evidence that functioned to weaken a claim, which was less than 25% even in the experimental condition. Of course, when interpreting these findings we should keep in mind that it would be impossible to expect all the idea units (100%) to include evidence. Students' contributions to the dialog may consist of claims or justifications, or explaining the relation between claims and evidence, which typically do not require evidence.

Nevertheless, these findings reveal that the objectives of using evidence to support one’s own position and especially to weaken opponents’ positions are not easily achieved (Kuhn & Moore, in press). Although students show preference for empirical data when they are asked (Sandoval & Çam, 2011), they do not readily employ them when they argue. To increase evidence use in argumentation may require engagement in an evidence-focused dialogic argumentation over an extended period of time during which meta-level understanding of argumentation as well as epistemological understanding would also develop (Kuhn, 2010; Kuhn, Iordanou, et al., 2008; Osborne et al., 2013; Ryu &Sandoval, 2012). Incorporating some additional reflective activities in the design of the *Socrates* LE, to prompt students to reflect on their overall use of evidence in argumentation when they analyze transcripts of their dialog, in addition to focusing on the use of evidence in particular arguments, might contribute to increasing the use of evidence in argumentation. In addition, asking students to explicitly provide their judgments of the arguments and counterarguments they produced, in the context of reflective activities may further support students’ meta-level understanding of the role of evidence in argumentation and result in greater use of evidence in argumentation.

The findings of the comparison group, who showed limited improvements in scientific argumentation, confirmed earlier findings in research in showing that curricula that do not focus on supporting the development of argumentation skills ultimately do not build students' skills in argumentation, particularly their ability to use evidence to back their arguments (Jiménez-Aleixandre, et al., 2000). The comparison group students’ goal was to explain the causes of Climate Change, whereas the experimental group students’ goal was to persuade through argumentation that Climate Change was due to a particular cause, either anthropogenic or natural. This difference in goals between the two conditions might have contributed to the difference observed between experimental and control-condition students in using evidence in argumentation at the final assessment. Trying to persuade a peer holding an opposing position, makes the need for using evidence more explicit, as compared to trying to develop a written explanation about a phenomenon. This finding is in line with Berland and Reiser's (2011) finding of enhanced engagement in evaluating and critiquing others' position when students were arguing with the goal to persuade rather than to make sense. In addition, note that although students in the comparison condition had some limited experience in dialogic argumentation, through their engagement in a dialog with a student holding an opposing position at initial assessment, they didn't exhibit any improvements in the use of evidence in argumentation. This finding suggests that limited experience in argumentation is not sufficient to induce development of argumentation skills.

Another implication of the comparison condition's performance is the role of content knowledge in scientific argumentation. Taking into consideration the fact that both comparison and experimental-condition students had access to the same content knowledge for the same time—through access to the same data base hosted in the *Socrates* learning environment designed for the present study—but only experimental-condition students increased the use of evidence in their dialogs has important theoretical and educational implications regarding the role of content knowledge in producing evidence-based arguments. Our findings extend previous research which showed that an adequate level of topic knowledge is required for students to engage in high quality argumentation (Von Aufschnaiter, Erduran, Osborne, & Simon, 2008), showing that possession of topic knowledge is probably a necessary but not a sufficient condition for skilled scientific argumentation. Comparison students did not exhibit any improvements in the use of evidence in their dialogs after studying relevant data on the topic; in addition, experimental-condition students exhibited transfer of their ability to use evidence when they argued on another socio-scientific topic, the transfer topic, for which they didn't have rich topic knowledge, besides some data that were provided in the scenario. This reveals that something else, beyond topic knowledge, needs to develop to support students' ability in using evidence in argumentation.

Supporting the development of argumentation norms, particularly in regard to employment of evidence in arguments, is not an easily achieved objective (Osborne et al., 2013). Only when students develop an understanding of persuasion as the ultimate goal of argumentation and an epistemological understanding that enables them to view scientific knowledge not as self-evident, absolute truth, but as the product of coordination of theory with evidence, will they be able to see the point of employing evidence in their arguments and only then will they have the motivation to use evidence to support their arguments and the critiques they offer to the opponents' arguments (Berland & Reiser, 2009). Previous research highlighted the role of teacher (Ryu & Sandoval, 2012) and of extended engagement in year-long interventions for developing norms of argumentation (Kuhn, Goh, et al., 2008; Kuhn et al., 2013). The present study adds to our understanding of argumentation by suggesting that a web-based learning environment that supports engagement in goal-based, sustained dialogic practice in the social context and reflection on the use of evidence in one's own dialogs, appears a promising pathway for developing students' norms of argumentation and the effective application of those norms in dialogic argumentation, particularly in respect to the ability of employing evidence in argumentation.

Acknowledgment

The research leading to these results has received funding from the European Union [European Communities Research Directorate General] Seventh Framework Programme (FP7/2007-2013) as part of project CoReflect under grant agreement n^{o} 217792.

References

Asterhan, C. S., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*(3), 626-639.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education, 2*(8), 797-817.

Berland, L., & Reiser, B. (2009).Making sense of argumentation and explanation. *Science Education*, *93*, 26-55.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191-216.

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, United Kingdom: Cambridge University Press.

Cardinal, R. N., & Aitken, M. R. (2013). *ANOVA for the behavioral sciences researcher*. Psychology Press.

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141-167.

Clark, D. B., Stegmann, K., Weinberger, A., Menekse, M., & Erkens, G. (2008). Technology enhanced learning environments to support students' argumentation. In S. Erduran &M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 217–243). Dordrecht, The Netherlands: Springer.

Duschl, R. A. (2008a). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based Research* (pp. 137–157). Dordrecht, The Netherlands: Springer.

Duschl, R. A. (2008b). Science education in 3 part harmony: Balancing conceptual, epistemic and social goals. *Review of Research in Education, 32,* 268-291.

Erduran, S., & Jiménez-Aleixandre, M. P. (Eds.). (2008). *Argumentation in science education: Perspectives from classroom-based research.* Dordrecht, The Netherlands: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument argument pattern for studying science discourse. *Science Education*, *88*, 915-933.

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*, 35-52.

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching, 38 (6)*, 663-687.

Iordanou, K. (2010). Developing argument skills across scientific and social domains. *Journal of Cognition and Development, 11*(3), 293-327.

Iordanou, K. (2013). Developing face-to-face argumentation skills: Does arguing on the computer help? *Journal of Cognition and Development, 14*(2), 292-320.

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or" doing science": Argument in high school genetics. *Science Education, 84(6),* 757-792.

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., ... Ulvik, M. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education, 90(4),* 632-655.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. The MIT Press.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education, 86(3),* 314-342.

Kuhn, D. (1991). The skills of argument. New York, NY: Cambridge University Press.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6,* 133-139.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94(5),* 810-824.

Kuhn, D., Goh,W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development, 79*(5), 1310 – 1328.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23*(4), 435-451.

Kuhn, D., & Moore, W. (in press). Argument as core curriculum. *Learning: Research and practice.*

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15,* 287-315.

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, Epistemological, and social dimensions of developing argumentive competence. *Cognition and Instruction, 31 (4),* 456-496.

Kyza, E. A., & Constantinou, C. P. (2007). STOCHASMOS: a web-based platform for reflective, inquiry-based teaching and learning. [Computer software]. Cyprus: Learning in Science Group.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*(11), 1387-1408.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492-509.

McNeill, K. L. (2009), Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education, 93*, 233-268.

McNeill, K. L., Lizotte, D. J, Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15*(2), 153-191.

McNeill, K. L., Pimentel, D. S., & Strauss, E. G. (2013) The impact of high school science teachers' beliefs, curricular enactments and experience on student learning during an inquiry-based urban ecology curriculum. *International Journal of Science Education, 35*(15), 2608-2644.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*(15), 1977-1999.

Osborne, G., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, *50*(3), 315-347.

Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education, 96*(3), 488-526.

Sadler, T. D., Chambers, W. F., & Zeidler, D. L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*,*26*(4),387-409.

Sadler, T. D., & Zeidler, D. L. (2005). Knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *International Journal of Science Education*, *89*, 71-93.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448-484.

Sandoval,W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89*, 634-656.

Sandoval, W. A. (2014). Science education's need for a theory of epistemological development. *Science Education*, *98*(3), 383-387.

Sandoval, W. A., & Çam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education*, *95*(3), 383-408.

Sandoval,W. A.,& Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic supports for science inquiry. *Science Education, 88*, 345-372.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, *45*(1), 101-131.

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation, 3,* 169-184.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35-62.

Table 1

Examples of Categories of Idea Units Containing Evidence Based on Evidence's Function

Supporting own claims	*The Argentine scientist GB estimated that a large proportion of greenhouse gas emissions in his country can be methane which is emitted by the 55 million cattle that graze his country plains.*
Weakening opponents' claims Weakenopp_C Weakenopp_A	S1: Nature is also responsible for climate change *S2: In graph 2 you can see that an increase in CO2 can be traced in 1800, at the same time that (Earth's) temperature increases. At that time also the industrial revolution began.* S3: The constant increase in the amount of CO2 in the atmosphere has resulted in global warming. *S4: This does not mean that only carbon dioxide is responsible for the phenomenon, there are also other gases such as nitric oxide, which are in a higher content than carbon dioxide, that also has a negative effect and contributes to the increase in the temperature.*

Meta-level talk about evidence	*Your claim could be more persuasive if… you had proofs regarding how nature is involved in our topic.*

Table 2

Percentage (and Standard Deviations) of Experimental Condition Participants' Idea Units Containing Evidence, Quantitative Evidence and Evidence which Functioned to Support own Claims or to Weaken Opponents' Claims on the Intervention Topic (Climate Change) by Time

	Initial Assessment	Final Assessment
Containing evidence	24.13% (8.62)	48.67% (8.21) *
Containing quantitative evidence	4.64% (5.06)	44.99% (9.49)**
Supporting own claims.	15.56% (8.08)	20.25% (11.48)
Weakening opponents' claims	5.31% (6.40)	23.33% (7.55)**
Weakenopp_C	3.21% (4.71)	18.75% (7.79)**
Weakenopp_A	2.10% (3.60)	4.50% (5.04)
Meta	3.17% (4.63)	4.90% (3.73)

*$p < .05$

**$p < .001$

Table 3

Percentage of Experimental Condition Participants' Idea Units Containing Accurate Data, Undocumented Evidence Claims from Personal Knowledge, Implicit and Explicit Reference to Evidence by Topic and Time

	Intervention Topic		Transfer Topic	
	Initial Assessment	Final Assessment	Initial Assessment	Final Assessment
Correct use of data from the LE	10%	59% **	38.8%	43.6%
Explicit reference to evidence from the LE	2.6%	52.9%	6.5%	21.1%*
Implicit reference to evidence	7.7%	5.7%	32.3%	22.5%
Undocumented evidence claims from personal knowledge	79%	29%**	56.5%	52.1%

*p <.05 McNemar test

** p < .001 McNemar test

Table 4

Percentage (and Standard Deviations) of Experimental Condition Participants' Idea Units Containing Evidence, Evidence which Functioned to Support Own Claims or to Weakening Opponents' Claims and Meta-*Level Talk about Evidence Over Time during the Intervention*

	Chat 1	Chat 2	Chat 3	Chat 4
Containing evidence	24.49% (12.04)	27.09% (5.01)	57.33% (16.23)*	64.69% (12.57)
Containing quantitative evidence	3.95% (4.63)	8.53% (9.60)	20.00% (11.58)	32.90% (9.70)
Supporting own claims	17.78% (9.77)	13.49% (6.42)	17.90% (10.09)	18.16% (11.02)
Weakening opponents' claims	7.60% (6.35)	10.35% (7.18)	27.60% (8.55)*	35.17% (10.28)
Weakenopp C	6.45% (5.41)	7.99% (6.14)	23.91% (7.13)*	29.32% (11.93)
Meta	3.12% (2.18)	3.24% (1.77)	11.82% (7.18)*	11.36% (9.66)

* $p < .05$

Table 5

Percentage (and Standard Deviations) of Experimental Condition Participants' Idea Units Containing Evidence, Quantitative Evidence and Evidence which Functioned to Support own Claims or to Weakening Opponents' Claims on the Transfer Topic (FGE) by Time

	Initial Assessment	Final Assessment
Containing evidence	26.75% (11.08)	39.83% (13.12)*
Containing quantitative evidence	4.42% (12.87)	12.87% (14.78)
Supporting own claims	13.56% (11.67)	15.22% (11.33)
Weakening opponents' claims	13.18% (14.52)	22.73% (14.77)
Weakenopp C	2.31% (5.53)	13.48% (7.31) *
Weakenopp A	10.87% (11.94)	9.25% (11.35)

$^*p < .05$

Table 6

Initial and Final Percentages (and Numbers) of Experimental Condition Participants who Produced Evidence, Idea Units Containing Overall_Weakenopp, Weakenopp_C and Support_Own Categories, by Topic and Time

Topic	Evidence		Overall_ Weakenopp		Weakenopp_C		Supportown	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
Intervention (CC)	50% (6)	100% (12)*	17% (2)	100% (12)*	17% (2)	100% (12)*	50%(6)	92% (11)
Transfer (FGE)	100% (12)	100% (12)	75% (9)	100% (12)	17% (2)	92% (11)*	67% (8)	83% (10)

*Significant change, $p < .05$, McNemar Test.

List of Figures

Appendix A

The scenario used for the transfer topic (FGE):

> Households of a new Town have to decide on the kind of fuel they will be using for producing electricity in their new Town. A problem has come up.
>
> A scientist, who was called to provide his viewpoint on the issue, claimed that Natural Gas should be used for producing electricity. "Natural gas is a gaseous mixture of hydrocarbons. Natural gas is colorless and odorless. The combustion of natural gas compared with that of other fuels such as coal or oil, has less harmful effects on the environment. For example, natural gas produces smaller quantities of carbon dioxide for every unit of energy produced than other sources, such as coal or oil. It is extracted from underground wells, where it is stored under high pressure. The gas has been formed in a way similar to the way of the formation of oil. The reserves of natural gas, although are limited at the moment compared to the reserves of coal, recently has increased due to the discovery of new sources of natural gas and in the future are likely to increase even more. The countries which have the biggest gas reserves are Russia, the United States of America, the United Arab Emirates, Canada and the United Kingdom."
>
> However, another scientist who was also called upon to provide his view about the issue claimed that Coal should be used for electricity production. "The Coal is a black rock used mainly as fuel. It is extracted from underground mines or open cavities in the ground. The countries with the biggest coal reserves are the United States, Russia, Australia, China, India and South Africa. Coal is the most economical option for electricity generation ─ its value corresponds to 1 / 10 of the price of other raw materials ─ and it is widely used in the industry because it contributes to the production of low cost industrial products. In addition, there are many coal reserves on Earth, which could last for many years. Although coal has been accused of releasing much more carbon dioxide than other materials when used for electricity production, scientists are recently studying ways for reducing the rates of carbon dioxide released during the process of electricity generation from coal."